A generalized uncertainty principle and sparse representation in pairs of bases

Article *in* IEEE Transactions on Information Theory · October 2002

DOI: 10.1109/TIT.2002.801410 · Source: IEEE Xplore

2 authors:

Michael Elad
Technion - Israel Institute of Technology
404 PUBLICATIONS **67,895** CITATIONS

SEE PROFILE

Alfred M. Bruckstein
Technion - Israel Institute of Technology
404 PUBLICATIONS **25,614** CITATIONS

SEE PROFILE

A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases

Michael Elad and Alfred M. Bruckstein

***Abstract*—An elementary proof of a basic uncertainty principle concerning pairs of representations of $\mathbb{R}^N$ vectors in different orthonormal bases is provided. The result, slightly stronger than stated before, has a direct impact on the uniqueness property of the sparse representation of such vectors using pairs of orthonormal bases as overcomplete dictionaries. The main contribution in this paper is the improvement of an important result due to Donoho and Huo concerning the replacement of the l_0 optimization problem by a linear programming (LP) minimization when searching for the unique sparse representation.**

***Index Terms*—Dictionaries comprising pairs of orthogonal bases, linear programming (LP), sparse representation, uncertainty principle.**

I. Introduction

Given a real column vector (also referred to as a signal) $\underline{S}$ in $\mathbb{R}^N$ it has a unique representation in every basis of this space. Indeed, if $\mathbf{\Phi} = \{\underline{\phi}_1, \underline{\phi}_2, \ldots \underline{\phi}_N\}$ are N orthogonal vectors of unit length, i.e., $\langle \underline{\phi}_i, \underline{\phi}_j \rangle = \delta_{ij}$, where the Kronecker symbol δ_{ij} equals 1 if $i = j$ and 0 if $i \neq j$, we have

$$\underline{S} = \left[\underline{\phi}_1 \underline{\phi}_2 \cdots \underline{\phi}_N\right] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \sum_{i=1}^{N} \alpha_i \underline{\phi}_i$$

where $[\underline{\phi}_1 \ \ \underline{\phi}_2 \ \ \cdots \ \ \underline{\phi}_N]$ is an N-by-N matrix, and the coefficients α_i are given by $\langle \underline{\phi}_i, \underline{S} \rangle$, since

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} \underline{\phi}_1^T \\ \underline{\phi}_2^T \\ \vdots \\ \underline{\phi}_N^T \end{bmatrix} \underline{S}.$$

Note that the inner product $\langle \underline{S}, \underline{U} \rangle$ may appear in our notation as $\underline{S}^T \underline{U}$ as well, following classic linear algebra conventions.

Suppose now that we have two different bases for $\mathbb{R}^N$: $\mathbf{\Phi} = \{\underline{\phi}_1, \underline{\phi}_2, \ldots \underline{\phi}_N\}$ and $\mathbf{\Psi} = \{\underline{\psi}_1, \underline{\psi}_2, \ldots \underline{\psi}_N\}$. Then every vector $\underline{S}$ has representations both in terms of $\{\underline{\phi}_i\}$ and in terms of $\{\underline{\psi}_i\}$. Let us write then

$$\underline{S} = \left[\underline{\phi}_1 \ \ \underline{\phi}_2 \ \ \cdots \ \ \underline{\phi}_N\right] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \left[\underline{\psi}_1 \ \ \underline{\psi}_2 \ \ \cdots \ \ \underline{\psi}_N\right] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}.$$

A question posed by Donoho and his coworkers is the following: is there some benefit in representing $\underline{S}$ in a joint, overcomplete set of vectors, say

$$\{\mathbf{\Phi}, \mathbf{\Psi}\} = \left\{\underline{\phi}_1, \underline{\phi}_2, \cdots \underline{\phi}_N, \underline{\psi}_1, \underline{\psi}_2, \cdots \underline{\psi}_N\right\}$$

(that can be called a "dictionary" concatenating the $\mathbf{\Phi}$ and $\mathbf{\Psi}$ bases)?

Indeed, we can consider a signal $\underline{S}^*$ that has a sparse representation in terms of the joint set of vectors forming the $\mathbf{\Phi}$ and the $\mathbf{\Psi}$ bases, but, in general, $\underline{S}^*$ will have highly nonsparse representations in either of these bases alone.

Sparse representations can have advantages in terms of compression of signals and/or in terms of understanding the underlying processes that generated them. The problem that arises, however, is that in terms of "dictionaries" of overcomplete set of vectors (as obtained by concatenating the basis vectors of $\mathbf{\Phi}$ and the $\mathbf{\Psi}$ to the N-by-$2N$ matrix $[\mathbf{\Phi} \ \ \mathbf{\Psi}]$) every signal has multiple representations. Of those multiple representations choosing one based on sparsity is a difficult optimization problem. Indeed, suppose we have

$$\underline{S} = \left[\underline{\phi}_1 \ \ \underline{\phi}_2 \ \ \cdots \ \ \underline{\phi}_N \ \ \underline{\psi}_1 \ \ \underline{\psi}_2 \ \ \cdots \ \ \underline{\psi}_N\right] \begin{bmatrix} \gamma_1^{\phi} \\ \gamma_2^{\phi} \\ \vdots \\ \gamma_N^{\phi} \\ \gamma_1^{\psi} \\ \gamma_2^{\psi} \\ \vdots \\ \gamma_N^{\psi} \end{bmatrix} = [\mathbf{\Phi} \ \ \mathbf{\Psi}] \underline{\gamma} = \sum_{i=1}^{N} \gamma_i^{\phi} \underline{\phi}_i + \sum_{i=1}^{N} \gamma_i^{\psi} \underline{\psi}_i.$$

Choosing $\underline{\gamma}$ involves solving an underdetermined set of N equations with $2N$ unknowns, and hence must be done subject to

Manuscript received December 6, 2001; revised March 26, 2002. This work emerged as a follow-up to David Donoho's extremely interesting talk (based on [4]) at the Technion–Israel Institute of Technology, Haifa, in December 2000. The material in this paper was presented in part at the International Conference on Image Processing (ICIP), Thessaloniki, Greece, November 2001, as an invited paper.

M. Elad is with the Department of Computer Science, SCCM Program, Stanford University, Stanford, CA 94304-9025 USA (e-mail: elad@sccm.stanford.edu).

A. M. Bruckstein is with the Department of Computer Science, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: freddy@cs.technion.ac.il).

Communicated by G. Battail, Associate Editor At Large.

Publisher Item Identifier 10.1109/TIT.2002.801410.

additional requirements on the solution. The additional requirement for sparsity would be to minimize the support of $\underline{\gamma}$, i.e., minimize the number of places where $\underline{\gamma}$ is nonzero. Hence we need to solve the problem

$$(P_0) \qquad \text{Minimize} \quad \|\underline{\gamma}\|_0 \quad \text{subject to} \quad \underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma}$$

where $\|\underline{\gamma}\|_0$ is the size of the support of $\underline{\gamma}$. This problem was addressed in two different approximation methods named "matching pursuit" [1], [2] and "basis pursuit" [3], [2]. Later, a result due to Donoho and Huo [4] formalized conditions under which the basis pursuit method exactly finds the desired sparse representation. They have shown that in case $\underline{S}$ has a "very" sparse representation, i.e., when there exists $\underline{\gamma}$ so that $\underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma}$ and $\|\underline{\gamma}\|_0 < f([\mathbf{\Phi} \ \ \mathbf{\Psi}])$, where $f(\cdot)$ is some given function to be specified later, then this sparse representation is the unique solution of not only (P_0) as defined above, but also of

$$(P_1) \qquad \text{Minimize} \quad \|\gamma\|_1 \quad \text{subject to} \quad \underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\gamma$$

where $\|\underline{\gamma}\|_1 = \sum_j |\gamma_j|$, the l_1-norm of $\underline{\gamma}$. This is an important result stating that "sparse" representations can be found by solving not a combinatorial search problem as implied by (P_0) but by solving the much simpler linear programming (LP) problem of minimizing the l_1-norm of $\underline{\gamma}$. The bound defining sparsity as provided by Donoho and Huo is

$$f([\mathbf{\Phi} \ \ \mathbf{\Psi}]) = \frac{1}{2}\left(1 + M^{-1}\right)$$

where

$$M = \sup_{1 \le i,j \le N} \left(\left|\left\langle \underline{\phi}_i, \underline{\psi}_j \right\rangle\right|\right).$$

The results of Donoho and Huo are based on exploiting an uncertainty principle stating that a signal cannot have representations that are simultaneously sparse in two orthonormal bases. This then leads to a result showing that sufficiently sparse representations in "dictionaries" concatenating the two bases must be unique, showing the uniqueness of the solution of problem (P_0) as defined above. Subsequently, for sufficiently sparse representations, it is shown that solving (P_1) leads also to the unique solution of (P_0).

In the general case discussed here Donoho and Huo showed in [4] that uniqueness of the solution of the (P_0) problem is ensured for $\|\underline{\gamma}\|_0 < \frac{1}{2}(1 + \frac{1}{M})$, and that in this case (P_1) provides this unique solution as well.

Here we follow the path of work of Donoho and Huo, and improve their bounds. First, we prove an "improved" uncertainty principle leading to better bounds yielding uniqueness of the (P_0) solution. The result is that uniqueness of the (P_0) solution is achieved for $\|\underline{\gamma}\|_0 < \frac{1}{M}$.

Our main contribution in this paper is an improvement of the result concerning the replacement of the (P_0) minimization problem with the (P_1) convex minimization, while achieving the same solution. We show that the solutions of the two problems (P_0) and (P_1) coincide for $\|\underline{\gamma}\|_0 < \frac{(\sqrt{2}-0.5)}{M}$. Doing so we enlarge the class of signals for which we can apply a simple, linear-programming based search for the optimal sparse representation.

II. The Basic Uncertainty Principle

We shall first prove a basic "uncertainty principle" concerning pairs of representations of a given vector $\underline{S}$ in two given orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$. Let us first state the result.

Theorem 1: Given a vector $\underline{S} \in \mathbb{R}^N$ and given two orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$, $\underline{S}$ may be represented both as $\underline{S} = \mathbf{\Phi}\alpha$ and $\underline{S} = \mathbf{\Psi}\beta$. For all such pairs of representations we have

$$\frac{\|\alpha\|_0 + \|\beta\|_0}{2} \ge \sqrt{\|\alpha\|_0 \cdot \|\beta\|_0} \ge \frac{1}{M}$$

where M is a scalar value denoting the maximal absolute value of cross-inner-products between vectors in $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

To prove this result, suppose a signal $\underline{S}$ has the representations

$$\underline{S} = \left[\underline{\phi}_1 \ \ \underline{\phi}_2 \ \ \cdots \ \ \underline{\phi}_N\right]\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix} = \mathbf{\Phi}\underline{\alpha}$$

$$= \left[\underline{\psi}_1 \ \ \underline{\psi}_2 \ \ \cdots \ \ \underline{\psi}_N\right]\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_N\end{bmatrix} = \mathbf{\Psi}\underline{\beta}$$

and without loss of generality assume that $\underline{S}^T\underline{S} = 1$, i.e., we have normalized the l_2 energy of the signal to 1. We have

$$\alpha_i = \langle \underline{S}, \underline{\phi}_i \rangle, \qquad \beta_j = \langle \underline{S}, \underline{\psi}_j \rangle.$$

Now let us write

$$\begin{aligned} 1 &= \underline{S}^T\underline{S} \\ &= [\alpha_1 \ \ \alpha_2 \ \ \cdots \ \ \alpha_N]\begin{bmatrix}\underline{\phi}_1^T\\ \underline{\phi}_2^T\\ \vdots\\ \underline{\phi}_N^T\end{bmatrix}\left[\underline{\psi}_1 \ \ \underline{\psi}_2 \ \ \cdots \ \ \underline{\psi}_N\right]\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_N\end{bmatrix} \\ &= \underline{\alpha}^T\mathbf{\Phi}^T\mathbf{\Psi}\underline{\beta} \end{aligned}$$

hence,

$$\begin{aligned} 1 &= [\alpha_1 \ \ \alpha_2 \ \ \cdots \ \ \alpha_N]\begin{bmatrix}\underline{\phi}_1^T\underline{\psi}_1 & \underline{\phi}_1^T\underline{\psi}_2 & \cdots & \underline{\phi}_1^T\underline{\psi}_N\\ \underline{\phi}_2^T\underline{\psi}_1 & \underline{\phi}_2^T\underline{\psi}_2 & \cdots & \underline{\phi}_2^T\underline{\psi}_N\\ \cdot & \cdot & & \cdot\\ \underline{\phi}_N^T\underline{\psi}_1 & \underline{\phi}_N^T\underline{\psi}_2 & \cdots & \underline{\phi}_N^T\underline{\psi}_N\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_N\end{bmatrix} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\langle\underline{\phi}_i, \underline{\psi}_j\rangle\beta_j. \end{aligned} \tag{1}$$

Writing

$$M = \sup_{i,j}(|\langle\underline{\phi}_i, \underline{\psi}_j\rangle|) = \sup\left(\left|\underline{\phi}_i^T\underline{\psi}_j\right|\right)$$

all the entries in the matrix $\mathbf{\Phi}^T\mathbf{\Psi}$ are smaller than M. We further have the (Parseval) energy preservation property

$$1 = \underline{S}^T\underline{S} = \sum \alpha_i^2 = \sum \beta_i^2.$$

Now assume that

$$\|\underline{\alpha}\|_0 = A \quad \text{and} \quad \|\underline{\beta}\|_0 = B.$$

Then (1) becomes

$$1 = \sum_{i'=1}^{A} \sum_{j'=1}^{B} \alpha_{i'} \langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle \beta_{j'}$$

where i' runs over the support of $\underline{\alpha}$ and j' runs over the support of $\underline{\beta}$. Now we can write

$$\begin{aligned} 1 &= \sum_{i'=1}^{A} \sum_{j'=1}^{B} \alpha_{i'} \langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle \beta_{j'} \\ &\le \sum_{i'=1}^{A} \sum_{j'=1}^{B} |\alpha_{i'}| \, |\langle \underline{\phi}_{i'}, \underline{\psi}_{j'} \rangle| \, |\beta_{j'}| \\ &\le \sum_{i'=1}^{A} \sum_{j'=1}^{B} |\alpha_{i'}| \cdot M \cdot |\beta_{j'}| \\ &= M \left(\sum_{i'=1}^{A} |\alpha_{i'}| \right) \left(\sum_{j'=1}^{B} |\beta_{j'}| \right). \end{aligned} \tag{2}$$

Next, in order to upper-bound the above expression, let us solve the following problem:

$$\text{maximize} \quad \left(\sum_{i=1}^{A} \alpha_i \right) \left(\sum_{j=1}^{B} \beta_j \right)$$

$$\text{subject to} \quad \alpha_i, \beta_j > 0, \quad \sum_{i=1}^{A} \alpha_i^2 = 1, \quad \sum_{j=1}^{B} \beta_j^2 = 1.$$

Since this problem is separable, we have to maximize both $(\sum_{i=1}^{A} \alpha_i)$ subject to $\alpha_i > 0$ and $\sum_{i=1}^{A} \alpha_i^2 = 1$, and $(\sum_{j=1}^{B} \beta_j)$ subject to $\beta_j > 0$ and $\sum_{j=1}^{B} \beta_j^2 = 1$. To solve these problems let us consider the following Lagrangian (note that we do not explicitly enforce the positivity constraint, but verify its correctness at the end of the process)

$$L(\underline{\alpha}) = \sum_{i=1}^{A} \alpha_i + \lambda \left[1 - \sum_{i=1}^{A} \alpha_i^2 \right],$$

$$\frac{\partial L}{\partial \alpha_i} = 1 - 2 \cdot \lambda \cdot \alpha_i = 0 \;\Rightarrow\; \alpha_i = \frac{1}{2\lambda}.$$

Next, using the unit-norm constraint we get

$$\begin{aligned} \sum_{i=1}^{A} \alpha_i^2 &= \sum_{i=1}^{A} \left(\frac{1}{2\lambda} \right)^2 \\ &= \frac{A}{4\lambda^2} = 1 \;\Rightarrow\; \lambda = \frac{\sqrt{A}}{2} \;\Rightarrow\; \alpha_i = \frac{1}{\sqrt{A}}. \end{aligned}$$

Therefore, the maximal value of $\sum_{i'=1}^{A} \alpha_{i'}$ equals $\frac{A}{\sqrt{A}} = \sqrt{A}$ and the maximal value of $\sum_{i'=1}^{A} \alpha_{i'} \sum_{j'=1}^{B} \beta_{j'}$ is $\sqrt{AB}$. Returning to our derivation of the uncertainty relations in (2) we now have

$$1 \le M \cdot \sum_{i'=1}^{A} \sum_{j'=1}^{B} |\alpha_{i'}| \, |\beta_{j'}| \le M \sqrt{AB}.$$

We have, therefore, obtained the result stated in Theorem 1: If we have two representations of a signal $\underline{S}$ in the bases $\mathbf{\Phi}[\underline{\phi}_1 \;\; \underline{\phi}_2 \;\; \cdots \;\; \underline{\phi}_N]$ and $\mathbf{\Psi}[\underline{\psi}_1 \;\; \underline{\psi}_2 \;\; \cdots \;\; \underline{\psi}_N]$ and the coefficient vectors $\underline{\alpha}$ and $\underline{\beta}$ have supports of sizes $\|\underline{\alpha}\|_0 = A$ and $\|\underline{\beta}\|_0 = B$ then

$$\sqrt{AB} \ge \frac{1}{M}$$

where $M = \text{Sup}\{|\langle \phi_i, \psi_j \rangle|, \forall (i,j)\}$. Using the well-known inequality between the geometric and arithmetic means, $\frac{A+B}{2} \ge \sqrt{AB}$, we also have that

$$\frac{A+B}{2} \ge \left(\sqrt{AB} \right) \ge \frac{1}{M} \;\Rightarrow\; A + B \ge \frac{2}{M}.$$

Donoho and Huo obtained, by emulating arguments for the sinusoidal and spike bases, a weaker result stating that

$$A + B \ge \left(1 + \frac{1}{M} \right)$$

(with respect to [4, eq. (7.1)]). Clearly

$$\frac{2}{M} = \frac{1}{M} + \frac{1}{M} \ge 1 + \frac{1}{M}$$

since $M \le 1$. In [4] it is said that their "… *general bound (i.e., (7.1)) can be a factor of two away from sharpness with respect to those (earlier particular, i.e., sinusoidal and spike bases) cases. Its generality can be an advantage in some cases.*"

We see here that an elementary derivation provides tight bounds that are not by a factor of two away from sharpness. Indeed, as $M \to \frac{1}{\sqrt{N}}$ (as it happens for spikes and complex sinusoids) $(1 + \frac{1}{M})$ goes to $1 + \sqrt{N}$ while $\frac{2}{M}$ goes to $2\sqrt{N}$, which is the bound claimed for this particular case (see, e.g., [4, eq. (2.1)] or [5, eq. (2.2)]).

In fact, the uncertainty result that we have obtained is even stronger since we got $\sqrt{AB} \ge \frac{1}{M}$. The use of the arithmetic mean instead of the geometric one looses tightness except for $A = B$. Also, this uncertainty result is of the form of the classical multiplicative uncertainty results (see, e.g., [2]) $\sigma_t^2 \cdot \sigma_\omega^2 \ge \frac{1}{4}$.

The value of M is crucial in the above arguments. For any pair of orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$ of $\mathbb{R}^N$ we have that $M \ge \frac{1}{\sqrt{N}}$. To see this, one simply notices that $\mathbf{\Phi}^T \mathbf{\Psi}$ is an orthonormal matrix, having the sum of squares of its entries equal to N. All entries cannot therefore be less than $1/\sqrt{N}$ since then we would have that the sum of all squared entries is less than N.

III. Uniqueness of Sparse Representations

A direct consequence of the uncertainty relation derived above is the following fact: if we have a "sparse" representation in terms of a dictionary that is the concatenation of two orthonormal bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$, it is unique. How sparse the representation should be to achieve such uniqueness depends crucially on the bound provided by the uncertainty principle. We first state the result as follows.

Theorem 2: If the signal $\underline{S}$ is to be represented using the concatenated dictionary $[\mathbf{\Phi}, \mathbf{\Psi}]$ ($2N$ vectors), for any two feasible representations denoted by $\underline{\gamma}_1, \underline{\gamma}_2 \in \mathbb{R}^{2N}$, we have

$$\|\underline{\gamma}_1\|_0 + \|\underline{\gamma}_2\|_0 \ge \frac{2}{M}.$$

Thus, there cannot be two different representations for the same vector having each less than $1/M$ nonzero entries, i.e., for any given representation $\underline{\gamma}$ we have that uniqueness is ensured by

$$\|\underline{\gamma}\|_0 < \frac{1}{M}.$$

This result derives easily from the following line of argumentation (taken from [4]). Suppose $\underline{\gamma}_1$ and $\underline{\gamma}_2$ are the coefficient vectors of two different representations of the same signal $\underline{S}$ in the dictionary $[\mathbf{\Phi} \ \ \mathbf{\Psi}]$, i.e.,

$$\underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma}_1 = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma}_2.$$

Then clearly

$$[\mathbf{\Phi} \ \ \mathbf{\Psi}][\underline{\gamma}_1 - \underline{\gamma}_2] = \underline{0}$$

or

$$\mathbf{\Phi} \cdot [\underline{\gamma}_1 - \underline{\gamma}_2]^{\phi} + \mathbf{\Psi} \cdot [\underline{\gamma}_1 - \underline{\gamma}_2]^{\psi} = \underline{0}$$

or

$$\mathbf{\Phi}\underline{\gamma}_{\Delta}^{\phi} = -\mathbf{\Psi}\underline{\gamma}_{\Delta}^{\psi} = \underline{\Lambda}.$$

Hence in this case, we have two vectors $\underline{\gamma}_{\Delta}^{\phi}$ and $\underline{\gamma}_{\Delta}^{\psi}$ defined as the "upper" and "lower" N values in $\underline{\gamma}_1 - \underline{\gamma}_2$, respectively. These two vectors are nonzero since they represent the same vector $\underline{\Lambda}$ in the two orthogonal bases.

Now the basic uncertainty principle states that if

$$\|\underline{\gamma}_{\Delta}^{\phi}\|_0 = A \quad \text{and} \quad \|\underline{\gamma}_{\Delta}^{\psi}\|_0 = B$$

then we must have

$$A + B \geq \left(2\sqrt{AB} \geq\right) \frac{2}{M}.$$

Suppose that the original representations were both sparse, i.e., that

$$\|\underline{\gamma}_1\|_0 < F \quad \text{and} \quad \|\underline{\gamma}_2\|_0 < F.$$

Then we must necessarily have

$$\|\underline{\gamma}_1 - \underline{\gamma}_2\|_0 < \|\underline{\gamma}_1\|_0 + \|\underline{\gamma}_2\|_0 < 2F.$$

On the other hand, we have

$$\|\underline{\gamma}_1 - \underline{\gamma}_2\|_0 = \left\|\underline{\gamma}_{\Delta}^{\phi}\right\|_0 + \left\|\underline{\gamma}_{\Delta}^{\psi}\right\|_0 = A + B.$$

Hence, sparsity of both $\underline{\gamma}_1$ and $\underline{\gamma}_2$ with bound F implies that

$$A + B < 2F.$$

But by the uncertainty principle we have

$$A + B \geq \frac{2}{M}.$$

In conclusion, if F would be M^{-1} or smaller, we would contradict the uncertainty principle if we would assume two different sparse representations. Hence, we have proved the uniqueness Theorem 2: If a signal $\underline{S}$ has a sparse representation in the dictionary $[\mathbf{\Phi} \ \ \mathbf{\Psi}]$ so that

$$\underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma} \quad \text{and} \quad \|\underline{\gamma}\|_0 < \frac{1}{M}$$

then this sparse representation is necessarily unique (i.e., there cannot be two different $\underline{\gamma}$'s obeying $\|\underline{\gamma}_0\| < M^{-1}$ that represent the same signal).

The bound $1/M$ is better than the one implied by the uncertainty principle stated in [4], which would yield $\frac{1}{2}(1 + M^{-1})$. This means that the uniqueness result will be true for much less sparse representations than those required by the bound provided in [4].

IV. Finding Sparse Representations Via l_1 Optimization

The next question that naturally arises is as follows: if a signal $\underline{S}$ has a sparse representation in a dictionary $[\mathbf{\Phi} \ \ \mathbf{\Psi}]$, how should we find it? Solving the l_0 optimization problem defined as

$$(P_0) \quad \text{Minimize } \|\underline{\gamma}\|_0 = \sum_{k=0}^{2N} \gamma_k^0 \quad \text{subject to } \underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma}$$

involves an unfeasible search (note that in the definition of the l_0 norm we define 0^0 to be 0). However, it was discovered experimentally that solving the l_1 optimization problem

$$(P_1) \quad \text{Minimize } \|\underline{\gamma}\|_1 = \sum_{k=0}^{2N} |\gamma_k| \quad \text{subject to } \underline{S} = [\mathbf{\Phi} \ \mathbf{\Psi}]\underline{\gamma}$$

often provides the sparse representation [3]. Donoho and Huo proved in [4] that the stronger sparsity condition

$$\|\underline{\gamma}\|_0 < \frac{1}{2}(1 + M^{-1})$$

ensures that the solution of the problem (P_1) yields the sparse solution of (P_0) too. This is a wonderful result, since (P_1) is essentially a LP problem!

To show that (P_1) provides also the (P_0) solution one has to prove that if

$$\|\underline{\gamma}\|_0 < F \quad \text{and} \quad [\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma} = \underline{S}$$

then, if there exists some other representation $[\mathbf{\Phi} \ \ \mathbf{\Psi}]\tilde{\underline{\gamma}} = \underline{S}$, we must have

$$\|\tilde{\underline{\gamma}}\|_1 \geq \|\underline{\gamma}\|_1.$$

In words, we need to show that the (unique) sparse $\underline{\gamma}$ is also "shortest" in the l_1 metric. As mentioned earlier, Donoho and Huo have shown that for $F = \frac{1}{2}(1 + M^{-1})$ this requirement is met. In what follows, we first describe the derivation of this result, then, in the next section we show how this bound can be improved.

Following [4] we have that if

$$[\mathbf{\Phi} \ \ \mathbf{\Psi}]\underline{\gamma} = \underline{S} = [\mathbf{\Phi} \ \ \mathbf{\Psi}]\tilde{\underline{\gamma}}$$

then

$$[\mathbf{\Phi} \ \ \mathbf{\Psi}][\tilde{\underline{\gamma}} - \underline{\gamma}] = 0.$$

Therefore, the difference vector $\underline{x} = \tilde{\underline{\gamma}} - \underline{\gamma}$ satisfies

$$\mathbf{\Phi}\underline{x}^{\phi} = \mathbf{\Psi}(-\underline{x}^{\psi}) \tag{3}$$

where we define the two vectors $\underline{x}^{\phi}$ and $\underline{x}^{\psi}$ as the "upper" and "lower" N values in $\underline{x} = \tilde{\underline{\gamma}} - \underline{\gamma}$, respectively. We need to show that for every nonzero vector $\underline{x}$ that obeys (3) we shall have

$$\sum_{k=1}^{2N} |\gamma_k + x_k| - \sum_{k=1}^{2N} |\gamma_k| \geq 0.$$

Hence, we need to show that

$$\sum_{\text{off support of } \underline{\gamma}} |x_k| + \sum_{\text{on support of } \underline{\gamma}} (|\gamma_k + x_k| - |\gamma_k|) \geq 0.$$

In order to shorten notations in the following analysis, we define

$$\sum_{\text{all}} |x_k| = \sum_{k=1}^{2N} |x_k|$$
$$\sum_{\text{off}} |x_k| = \sum_{\text{off support of } \underline{\gamma}} |x_k|$$
$$\sum_{\text{on}} |x_k| = \sum_{\text{on support of } \underline{\gamma}} |x_k|.$$

The inequality $|v+m| \geq |v|-|m|$ implies $|v+m|-|v| \geq -|m|$, so we have

$$\sum_{\text{off}} |x_k| + \sum_{\text{on}} (|\gamma_k + x_k| - |\gamma_k|) \geq \sum_{\text{off}} |x_k| - \sum_{\text{on}} |x_k|.$$

Putting this into the preceding inequality we get

$$\sum_{\text{off}} |x_k| - \sum_{\text{on}} |x_k| \geq 0.$$

Adding $2\sum_{\text{on}} |x_k|$ to both sides results in

$$\sum_{\text{off}} |x_k| + \sum_{\text{on}} |x_k| = \sum_{\text{all}} |x_k| \geq 2 \sum_{\text{on}} |x_k|$$

which finally gives

$$\frac{1}{2} \sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k| \geq 0 \text{ or equivalently } \frac{\sum_{\text{on}} |x_k|}{\sum_{\text{all}} |x_k|} \leq \frac{1}{2}. \quad (4)$$

Therefore, if we shall show that for some value of the support size of $\underline{\gamma}$ (denoted by F), all $\underline{x}$ obeying (3) also satisfy (4), then the assertion $||\tilde{\underline{\gamma}}||_1 > ||\underline{\gamma}||_1$ follows, and the (P_1) problem can replace the (P_0) problem while searching for the sparse $\underline{\gamma}$. Following [4], in order to prove (4), consider the following problem:

$$\text{minimize} \quad \sum_{i=1}^{N} |x_i^{\phi}| + \sum_{i=1}^{N} |x_i^{\psi}|$$
$$\text{subject to} \quad \left\{ x_k^{\phi/\psi} = V \text{ and } \mathbf{\Phi}[\underline{x}^{\phi}] = \mathbf{\Psi}[-\underline{x}^{\psi}] \right\}$$

where V is an arbitrary nonzero scalar, k is an arbitrary integer index in the range $[1, N]$, $x_k^{\phi/\psi}$ denotes either x_k^{ϕ} or x_k^{ψ}, and x_i^{ϕ}, x_i^{ψ} are the ith entries of the vectors $\underline{x}^{\phi}$ and $\underline{x}^{\psi}$, respectively ($\underline{x}^{\phi}$ and $\underline{x}^{\psi}$ where defined above—see (3)). This minimization problem consists of finding the smallest possible denominator in (4) while assuming that some arbitrary (kth) entry in either $\underline{x}^{\phi}$ or $\underline{x}^{\psi}$ is nonzero and equal to V. If the minimum attained for all conditions $x_k^{\phi/\psi} = V$ is given by some $f(V) \geq |V|\tilde{F}(M)$ then we shall be able to say that

$$\frac{\sum_{\text{on}} |x_i|}{\sum_{\text{all}} |x_j|} = \sum_{\text{on}} \frac{|x_i|}{\sum_{\text{all}} |x_j|} \leq \sum_{\text{on}} \frac{|V|}{|V| \cdot \tilde{F}(M)} = \frac{1}{\tilde{F}(M)} \cdot ||\underline{\gamma}||_0.$$

But we have assumed that $||\underline{\gamma}||_0 < F$, hence we shall have

$$\frac{\sum_{\text{on}} |x_i|}{\sum_{\text{all}} |x_i|} < \frac{F}{\tilde{F}(M)}.$$

If we have $F \leq \frac{1}{2}\tilde{F}(M)$ then the required condition for the optimality of $\underline{\gamma}$ for the l_1 problem will be met. We shall next prove that $\tilde{F}(M) \geq 1 + \frac{1}{M}$ and it will lead to the result that if $F \leq \frac{1}{2}(1 + \frac{1}{M})$, (4) holds as needed.

Since we have $\mathbf{\Phi} = \underline{x}^{\phi} = -\mathbf{\Psi}\underline{x}^{\psi}$ we can write $\underline{x}^{\phi} = -\mathbf{\Phi}^T\mathbf{\Psi}\underline{x}^{\psi}$. Suppose without loss of generality that the condition we have is $x_k^{\phi} = V$. Then we have

$$||\underline{x}^{\phi}||_{\infty} = \max_j ||x_j^{\phi}|| = |x_{j_0}^{\phi}| \text{ for some index } j_0$$
$$= \left| -\left[\mathbf{\Phi}^T\mathbf{\Psi}\right]_{\text{row}j_0} \cdot \underline{x}^{\psi} \right| \leq M \cdot ||\underline{x}^{\psi}||_1.$$

Hence,

$$||\underline{x}^{\psi}||_1 \geq \frac{||\underline{x}^{\phi}||_{\infty}}{M} \geq \frac{|V|}{M}.$$

We also obviously have

$$||\underline{x}^{\phi}||_1 \geq |V|$$

hence

$$||\underline{x}^{\psi}||_1 + ||\underline{x}^{\phi}||_1 \geq |V| \left(1 + \frac{1}{M}\right).$$

Therefore, we may take

$$\tilde{F}(M) = \left(1 + \frac{1}{M}\right).$$

This proves that if

$$||\underline{\gamma}||_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right)$$

the sparse representation will be provided by the solution of (P_1) as well.

So far we have seen that if

$$||\underline{\gamma}||_0 < \frac{1}{M}$$

the sparse representation is unique and if

$$||\underline{\gamma}||_0 < \frac{1}{2}\left(1 + \frac{1}{M}\right)$$

the unique sparse representation is provided by solving an LP problem (P_1). Hence, if

$$\frac{1}{2}\left(1 + \frac{1}{M}\right) \leq ||\underline{\gamma}||_0 < \frac{1}{M}$$

we have uniqueness of the (P_0) solution, but we cannot be sure that (P_1) will provide this solution. So, obviously, there remains a gap, and the natural question is what happens for sparsity of vectors in this gap. A wealth of simulations that we have performed has shown that for signals with (unique) sparse representation in this gap the equivalence between the (P_0) and the (P_1)

solutions remains true. Motivated by these empirical results, we succeeded to prove that the gap is indeed much narrower.

V. Improving the Bound for l_1 Optimization

In the preceding section we closely followed Donoho and Huo's work and reproduced their results. Next we take a different route in order to obtain an improved bound necessary for the equivalence between (P_0) and (P_1). The result to be established in this section states the following.

Theorem 3: If the signal $\underline{S}$ has a sparse representation $\underline{\gamma}$ (i.e., $\underline{S} = [\mathbf{\Phi} \quad \mathbf{\Psi}]\underline{\gamma}$) such that

$$||\underline{\gamma}||_0 < \frac{(\sqrt{2} - 0.5)}{M} = \frac{0.9142}{M}$$

then the l_1-norm minimization solution coincides with the minimization of the l_0 norm.

Let us prove this claim. Equations (3) and (4) can be re-interpreted as an optimization problem of the following form:

$$\begin{aligned} \text{minimize} \quad & \left[\tfrac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|\right] \\ \text{subject to} \quad & \mathbf{\Phi}\underline{x}^{\phi} = \mathbf{\Psi}(-\underline{x}^{\psi}). \end{aligned}$$

This problem should be solved for various values of $||\underline{\gamma}||_0$ and all profiles of nonzero entries in $\underline{\gamma}$. As $||\underline{\gamma}||_0$ increases the minimum of the above expression decreases and tends to a negative value. The largest $||\underline{\gamma}||_0$ that yields a minimum that is still above zero as a consequence of (4) will be the bound on the sparsity of $\underline{\gamma}$, ensuring equivalence between (P_0) and (P_1). However, working with the above minimization problem is complicated because of several reasons.

1) We have not explicitly set conditions on $\underline{x}$ to avoid the trivial solution $\underline{x} = \underline{0}$.
2) The problem involves both the entries of the vector $\underline{x}$ and their absolute values.
3) The orthonormal matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ appear explicitly, and we would like them to influence the result only through the parameter M they define.
4) The problem is clearly sensitive not only to the number of nonzero elements in the support of $\underline{\gamma}$, but also to their position.

In order to solve the first difficulty, we introduce an additional constraint that prevents the trivial solution. Such a constraint could be posed on the l_2 or the l_1 norm of the unknown vector $\underline{x}$, i.e., $\underline{x}^T\underline{x} = 1$ or $\sum_{\text{all}} |x_k| = 1$. The newly added constraint will clearly not change the sign of the result, i.e., if for some $||\underline{\gamma}||_0$ the minimization result becomes negative, then it would have been so without the constraint as well. Thus, the new constraint does not interfere with our goals in this optimization problem. As to the other difficulties, we solve them via the definition of an alternative minimization problem. In this new problem, we minimize the same function, but pose a weaker set of constraints as follows:

$$\begin{aligned} \text{minimize} \quad & \left[\frac{1}{2}\sum_{\text{all}} |x_k| - \sum_{\text{on}} |x_k|\right] \\ \text{subject to:} \quad & 1.\ \underline{X}_1 \le M \cdot \mathbf{1}_{N\times N}\underline{X}_2 \\ & 2.\ \underline{X}_2 \le M \cdot \mathbf{1}_{N\times N}\underline{X}_2 \\ & 3.\ \sum_{\text{all}} |x_i| = 1 \end{aligned}$$

where we define $\underline{X}_1$ and $\underline{X}_2$ as the absolute values of the original vectors $\underline{x}^{\phi}$ and $\underline{x}^{\psi}$, respectively, and where $\mathbf{1}_{N\times N}$ is an N-by-N matrix all the entries of which are ones.

The first and second constraints simply use the fact that any given entry in one of the vectors ($\underline{x}^{\psi}$ or $\underline{x}^{\phi}$) cannot be greater than M multiplying the sum of the absolute entries in the other vector. Clearly, every feasible solution of the original constraint set is also a feasible solution of the new constraint set, but not *vice versa*. Thus, if the minimum of the function is still positive using the new constraint set, it implies that it is surely positive using the original constraint set.

Looking closely at the newly defined optimization problem, we can further simplify it by noticing first that the sum over all the support that appears in the function is known to be 1 due to the third constraint. Also, only the absolute values of the unknown vector $\underline{x}$ play a role here. Exploiting these two properties we can rewrite the new problem and get

$$\begin{aligned} \text{Minimize} \quad & \left[\frac{1}{2} - \mathbf{1}_{\gamma_1}^T\underline{X}_1 - \mathbf{1}_{\gamma_2}^T\underline{X}_2\right] \\ \text{subject to:} \quad & 1.\ \underline{X}_1 \le M \cdot \mathbf{1}_{N\times N}\underline{X}_2 \\ & 2.\ \underline{X}_2 \le M \cdot \mathbf{1}_{N\times N}\underline{X}_1 \\ & 3.\ \mathbf{1}_N^T(\underline{X}_1 + \underline{X}_2) = 1 \\ & 4.\ \underline{X}_1 \ge 0, \quad \underline{X}_2 \ge 0 \end{aligned} \tag{5}$$

where we added the fourth constraint regarding positivity of the unknowns, i.e., the entries of $\underline{X}_1$ and $\underline{X}_2$. The notations $\mathbf{1}_{\gamma_1}$ and $\mathbf{1}_{\gamma_2}$ stand for vectors of length N, $[\mathbf{1}_{\gamma_1}\mathbf{1}_{\gamma_2}]$ being the $(2N)$-vector with ones where $\underline{\gamma}$ is nonzero and $\mathbf{1}_N$ is simply an N-vector containing all ones.

If we assume that there are K_1 nonzero components in $\mathbf{1}_{\gamma_1}$ and K_2 nonzero components in $\mathbf{1}_{\gamma_2}$, then $K_1 + K_2 = ||\gamma||_0$. In the new formulation, we can assume without loss of generality that the K_1 (K_2) nonzero components are located at the first positions of vector $\underline{X}_1$ $(\underline{X}_2)$, due to the symmetrical form of the constraints.

The problem we have obtained is a classical LP problem and, as such, has a unique local minimum point which is also the unique global minimum point. Let us bring it to its canonical form

$$(P) \quad \text{Minimize } \underline{C}^T\underline{Z} \quad \text{subject to } \boldsymbol{A}\underline{Z} \ge \underline{B}, \quad \underline{Z} \ge 0.$$

(P) denotes the fact that this is a primal LP form. The matrices $(\boldsymbol{A}, \underline{Z}, \underline{B}, \underline{C})$ involved are defined as follows:

$$\underline{Z}^T = \left[\underline{X}_1^T \;\; \underline{X}_2^T\right]$$
$$\underline{C}^T = \left[-\mathbf{1}_{\gamma_1}^T \;\; -\mathbf{1}_{\gamma_2}^T\right]$$
$$\underline{B}^T = \left[0 \cdot \mathbf{1}_N^T \;\; 0 \cdot \mathbf{1}_N^T \;\; 1 \;\; -1\right]$$
$$\boldsymbol{A} = \begin{bmatrix} -\boldsymbol{I}_N & M \cdot \mathbf{1}_{N\times N} \\ M \cdot \mathbf{1}_{N\times N} & -\boldsymbol{I}_N \\ \mathbf{1}_N^T & \mathbf{1}_N^T \\ -\mathbf{1}_N^T & -\mathbf{1}_N^T \end{bmatrix}.$$

Notice that the equality constraint in the original problem is replaced by two inequalities, in order to fit the required canonic form. In the optimization problem defined in (5) we wanted conditions so that the minimum will not be negative. After removing the constant $1/2$ from the function to be minimized, the new requirement becomes $\underline{C}^T\underline{Z} \geq -0.5$.

It is still difficult to give an analytic form to the solution of the LP problem we obtained. Instead, we shall exploit the dual LP problem of the form

$$(D) \quad \text{maximize } \underline{B}^T\underline{U} \quad \text{subject to } \boldsymbol{A}^T\underline{U} \leq \underline{C}, \quad \underline{U} \geq 0.$$

(D) here denotes that this is a dual LP form, with the same matrices as in the primal problem (see, e.g., [6]).

The approach we are going to take is as follows: We know that the primal and the dual problems yield the same optimal value [6], i.e.,

$$\text{minimize } \left\{\underline{C}^T\underline{Z}\right\} = \text{maximize } \left\{\underline{B}^T\underline{U}\right\}.$$

We require this optimal value to be higher than or equal to -0.5. In the dual problem, if we find a feasible solution $\underline{U}$ such that $\underline{B}^T\underline{U} \geq -0.5$, we guarantee that the maximal value is also above -0.5, and thus we fulfill the original requirement on the primal problem.

Let us consider the following parameterized form for a feasible solution for $\underline{U}$:

$$\underline{U}^T = [\mathbf{1}_{\gamma_1}^T \quad \alpha \cdot \mathbf{1}_{\gamma_2}^T \quad \beta \quad \delta]$$

where α, β, and δ are real scalar parameters. Using previous notations, there are K_1 nonzero components in $\mathbf{1}_{\gamma_1}$ and K_2 nonzero components in $\mathbf{1}_{\gamma_2}$. We assume without loss of generality that $K_1 \leq K_2$ (the problem is perfectly symmetric with respect to these two sections). We also assume that $0 \leq \alpha \leq 1$. This way, three parameters (α, β, δ) govern the entire solution $\underline{U}$, and we need to find requirements on them in order to guarantee that the proposed solution is indeed feasible.

Substituting the proposed solution form into the constraint inequalities of the dual problem we get

$$-\mathbf{1}_{\gamma_1} + \alpha M \cdot \mathbf{1}_{N\times N}\mathbf{1}_{\gamma_2} + (\beta - \delta)\mathbf{1}_N \leq -\mathbf{1}_{\gamma_1}$$
$$-\alpha\mathbf{1}_{\gamma_2} + M \cdot \mathbf{1}_{N\times N}\mathbf{1}_{\gamma_1} + (\beta - \delta)\mathbf{1}_N \leq -\mathbf{1}_{\gamma_2}$$

and rearranging these inequalities we get

$$\alpha M \cdot \mathbf{1}_{N\times N}\mathbf{1}_{\gamma_2} + (\beta - \delta)\mathbf{1}_N \leq 0$$
$$(1-\alpha)\mathbf{1}_{\gamma_2} + M \cdot \mathbf{1}_{N\times N}\mathbf{1}_{\gamma_1} + (\beta - \delta)\mathbf{1}_N \leq 0.$$

The multiplication $\mathbf{1}_{N\times N}\mathbf{1}_{\gamma_2}$ is simply $K_2\mathbf{1}_N$, and similarly, $\mathbf{1}_{N\times=N}\mathbf{1}_{\gamma_1} = K_1\mathbf{1}_N$. Thus, the first set of N inequalities actually reduces to the simple scalar requirement

$$\alpha M K_2 + (\beta - \delta) \leq 0.$$

If $0 \leq \alpha \leq 1$, as assumed earlier, the second set of inequalities can be replaced by the scalar requirement

$$(1-\alpha) + MK_1 + (\beta - \delta) \leq 0.$$

(For the $N - K_2$ places where $\mathbf{1}_{\gamma_2}$ has zero components, we get the inequality $MK_1 + (\beta - \delta) \leq 0$, which is weaker than the above one.)

Solving the two inequalities as equalities, we get

$$\alpha M K_2 = (1-\alpha) + MK_1 \;\Rightarrow\; \alpha = \frac{1+MK_1}{1+MK_2}.$$

We see that, indeed, our assumption $0 \leq \alpha \leq 1$ is correct since we assumed $K_1 \leq K_2$ and we have $M > 0$.

So far we have found an expression for the first parameter α. As to the other two, substituting α in the above equations we get

$$(\beta - \delta) = -MK_2 \cdot \frac{1+MK_1}{1+MK_2} < 0.$$

Thus, we can choose $\beta = 0$ and δ will be the above expression multiplied by -1. This way, we have satisfied all the inequality constraints, and obtained a solution $\underline{U}$ which is also nonnegative in all its entries.

Now that we have established that the proposed solution is feasible, let us look at the value of the function to be maximized. This function is simply $\underline{B}^T\underline{U} = (\beta - \delta)$. So we should require

$$\underline{B}^T\underline{U} = (\beta - \delta) = -MK_2 \cdot \frac{1+MK_1}{1+MK_2} \geq -\frac{1}{2}.$$

Thus,

$$MK_2 \cdot \frac{1+MK_1}{1+MK_2} \leq \frac{1}{2} \;\Rightarrow\; 2M^2K_1K_2 + MK_2 - 1 \leq 0. \quad (6)$$

We have, therefore, obtained a requirement on K_1 and K_2 which is posed in terms of the parameter M. Thus, given a vector $\underline{S} \in \mathbb{R}^N$, the two orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, and their induced cross-product factor M, we solve the (P_1) problem and obtain a candidate representation $\underline{\gamma}$. From this representation, we can determine K_1 and K_2, the respective numbers of nonzero components in the two N-sections of $\underline{\gamma}$. Putting them into the inequality (6) we immediately know whether this is also the (P_0) solution as well, and hence also the sparsest representation.

We now proceed and produce a simpler sparsity condition on the representation $\underline{\gamma}$ ensuring that (P_1) produces the unique and sparse (P_0) solution. Assuming that we know K_2, the requirement on K_1 is

$$2M^2K_1K_2 + MK_2 - 1 \leq 0 \;\Rightarrow\; K_1 \leq \frac{1}{M} \cdot \frac{1 - MK_2}{2MK_2}.$$

Adding K_2 to both sides of the above inequality we get

$$K_1 + K_2 \leq \frac{1}{M} \cdot \frac{1-MK_2}{2MK_2} + K_2$$
$$= \frac{1}{M} \cdot \frac{1 - MK_2 + 2(MK_2)^2}{2MK_2}.$$

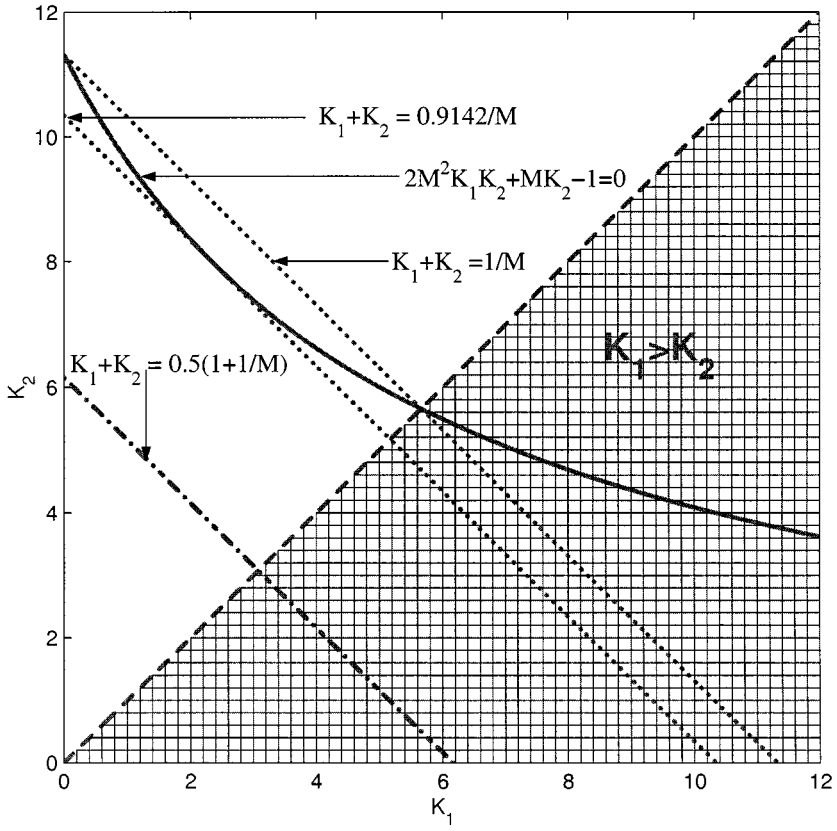

Fig. 1. A graphic description of the bounds obtained.

Let us look at the term which multiplies $1/M$ in the above bound. This term is a function of $x = MK_2$ which is clearly nonnegative. The minimum of this term is given by cancelling the derivative of the function $f(x) = (1 - x + 2x^2)/(2x)$, thus solving the equation $0 = 2x^2 - 1$, which results in $x = \pm\sqrt{0.5}$. The negative solution is irrelevant, and the positive one is indeed the minimum (as easily verified with the second derivative), where we get that $f(\sqrt{0.5}) = \sqrt{2} - 0.5 = 0.9142$. Thus, our final result is the requirement

$$\|\underline{\gamma}\|_0 = K_1 + K_2 \leq \left(\sqrt{2} - 0.5\right) \cdot \frac{1}{M}.$$

To summarize, we got that if $\|\underline{\gamma}\|_0 \leq (\sqrt{2} - 0.5)/M$ then the dual LP necessarily gives a value above -0.5, which in turn assures that the minimum of the primal problem is above -0.5 as well. Thus, it is guaranteed that the l_1 norm of the solution with this number of nonzero components is the smallest possible, so we have proved that the (P_1) problem yields the solution of (P_0) as well.

The result obtained is better than the $0.5(1 + 1/M)$ bound asserted by Donoho and Huo. As an example, for $M = 1/\sqrt{N}$, we get that for $N = 16$ the old (DH) requirement is to have less than 2.5 nonzero components, while we (EB) require 3.65 and below. For $N = 64$, the old requirement (DH) is 4.5 and below, while the new bound requires 7.3 or less nonzero components. As N goes to infinity the ratio between the two results becomes

$$\frac{\frac{(\sqrt{2}-0.5)}{M}}{0.5\left(1+\frac{1}{M}\right)} = \frac{\left(\sqrt{2}-0.5\right)\sqrt{N}}{0.5\left(1+\sqrt{N}\right)} \rightarrow 2\sqrt{2} - 1 = 1.8284.$$

Fig. 1 shows graphically how (6) compares to the alternatives

$$K_1 + K_2 \leq \frac{(\sqrt{2} - 0.5)}{M}$$

and

$$K_1 + K_2 \leq \frac{1}{M}.$$

In this graph, we have assumed $M = 1/\sqrt{128}$. Note that since $0 \leq K_1 \leq K_2$, only the upper left part of the graphs is relevant, and thus we have masked the nonrelevant zone. We can see that the $1/M$ bound (which is also the uniqueness bound) is valid at the extremes, whereas the $(\sqrt{2} - 0.5)/M$ bound is relevant in the middle of the K_1 zone.

Fig. 2 graphically describes how the various bounds developed by Donoho and Huo (and thus denoted by DH) compare with the new bounds established here (denoted EB).

One question that remains at the moment unanswered is whether the $(\sqrt{2} - 0.5)$ coefficient we got reflects the true behavior of the (P_0) problem versus the (P_1) one, or is it emerging due to the still loose bounding approximation in our proof. For example, if we take the trivial (spike) and Hadamard bases for $N = 128$, we get that $1/M = 1/\sqrt{128} = 1/11.31$. Hence, the uniqueness bound requires $\|\underline{\gamma}\|_0 \leq 1/M = 11.31$. On the other hand, the new bound found here, $(\sqrt{2} - 0.5)/M$, requires less than 10.34 nonzero entries, which in practice means 10 and below in order to ensure that the (P_1) and the (P_0) problems coincide. An experiment that would test this case with 11 nonzero entries and produce a counterexample where (P_0) and (P_1) lead to different solutions could be

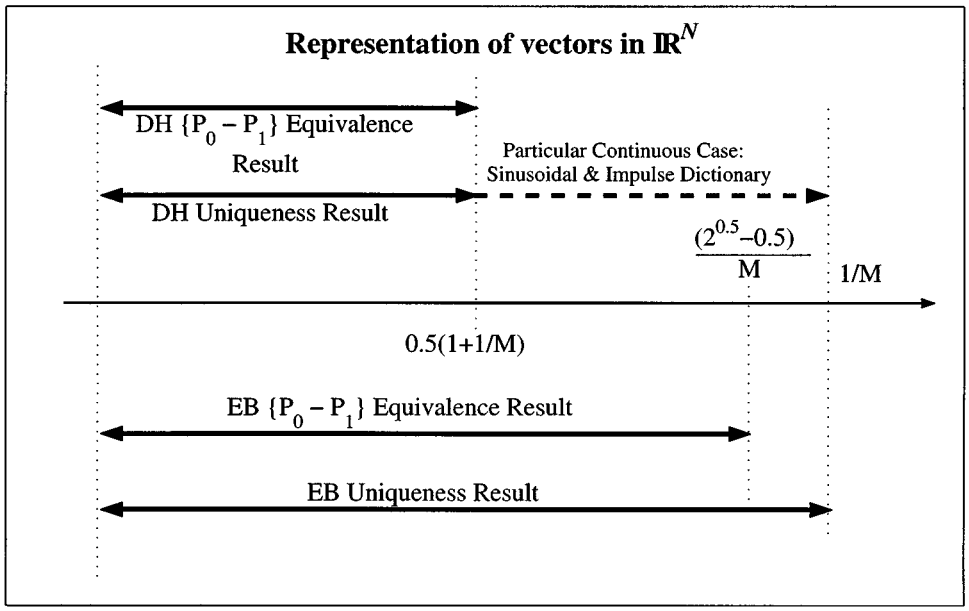

Fig. 2. A graphic illustration of the bounds due to Donoho and Huo (DH) versus the ones developed in this paper (EB).

beneficial. The search for such a counterexample is currently under way but so far we could not find any case like this.

It is interesting to note that if we assume that $K_1 = 0$, the requirement made in (6) leads to the requirement $K_2 = \|\underline{\gamma}\|_0 \leq \frac{1}{M}$. Similarly, if $K_2 = K_1$ then (6) leads to the requirement $2K_2 = \|\underline{\gamma}\|_0 \leq \frac{1}{M}$. Note that these two cases are the extremes in the range of K_1 $(0 \leq K_1 \leq K_2)$. Thus, for these two cases, the $1/M$ bound is valid and there is no gap with respect to the uniqueness bound.

Returning to the suggested search for a counterexample for the 11 nonzero components described above, it is clear now that choosing the nonzero components in one of the above two extremes will not lead to a successful hunt. Instead, one has to choose $K_2 \approx \frac{\sqrt{0.5}}{M} = 8$ and $K_1 = 11 - 8 = 3$ in order to maximize the chances for success. Again, simulations done so far have not found such a special case.

Another important question is how much of the entire range of signals in $\mathbb{R}^N$ are we covering by the sparsity condition we have formed? What is the "measure" of signals that require more than $(\sqrt{2} - 0.5)/M$ nonzero components in their sparse representation? These and more questions will be the subject of future investigations.

VI. Conclusion

Given a vector $\underline{S} \in \mathbb{R}^N$, and given an orthonormal N-by-N matrix $\mathbf{\Phi}$ representing a basis, we can uniquely represent $\underline{S}$ using this basis and get $\underline{\alpha} = \mathbf{\Phi}^T \underline{S}$. Assuming that we have a second orthonormal basis $\mathbf{\Psi}$ and a second representation $\underline{\beta} = \mathbf{\Psi}^T \underline{S}$, the first question we have addressed in this paper is whether there is a lower bound on the number of nonzero components in these two representations. It was found that

$$\frac{\|\alpha\|_0 + \|\beta\|_0}{2} \geq \sqrt{\|\alpha\|_0 \cdot \|\beta\|_0} \geq \frac{1}{M}$$

where M is a scalar value denoting the maximal absolute value of cross-inner-products between vectors in $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

The second question answered in this paper refers to the uniqueness of sparse representation with overcomplete dictionaries. Assuming that the signal $\underline{S}$ is to be represented using a dictionary that contains $\mathbf{\Phi}$ and $\mathbf{\Psi}$ ($2N$ vectors), it is clear that there are numerous possible representations. It was established that for any two feasible representations denoted by $\underline{\gamma}_1, \underline{\gamma}_2 \in \mathbb{R}^{2N}$, we have that

$$\|\underline{\gamma}_1\|_0 + \|\underline{\gamma}_2\|_0 \geq \frac{2}{M}.$$

Thus, there cannot be two different representations for the same vector having each less than $1/M$ nonzero entries, i.e., for any given representation $\underline{\gamma}$, we have that uniqueness is ensured by

$$\|\underline{\gamma}\|_0 < \frac{1}{M}.$$

The main contribution of the paper concentrated to the way to find the sparse representation over an overcomplete dictionary as described above. Finding the optimal representation by minimizing the l_0 norm is a highly complicated nonconvex combinatorial optimization search. An alternative approach based on l_1-norm minimization was proposed by Donoho and Huo and proved to lead to the same result for sufficiently sparse signals. Their result was that if the signal has a sparse representation with no more than $0.5(1 + 1/M)$ nonzero entries, minimization of the l_1 norm can replace the original l_0-norm minimization while ensuring the same result. Obviously, such a result is very valuable since l_1-norm minimization leads to an LP problem, which is a convex problem, while the l_0-norm minimization is a hard nonconvex problem. In this paper, we have improved this sparsity bound and found that if $\underline{S}$ has a sparse representation $\underline{\gamma}$ such that

$$\|\underline{\gamma}\|_0 < \frac{(\sqrt{2} - 0.5)}{M} = \frac{0.9142}{M}$$

then the l_1-norm minimization solution coincides with the minimization of the l_0 norm. In fact, the bound obtained in this paper

is better than $\frac{(\sqrt{2}-0.5)}{M}$, since we obtained the above requirement in terms of K_1 and K_2 (the number of nonzero components in the two N-sections of the representations) which obey

$$2M^2 \cdot K_1 K_2 + MK_2 - 1 \leq 0.$$

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor At Large, Dr. Gérard Battail and the referees for valuable comments that helped improve the clarity of presentation of this paper.

REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
[2] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA: Academic, 1998.
[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, Jan. 2001.
[4] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.
[5] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49/3, pp. 906–931, June 1989.
[6] D. Bertsekas, *Non-Linear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1995.